THE LINUX FOUNDATION | Research

オープンソースとAIの未来

エージェントがシステム、先例、そしてソフトウェアにおける人間の役割にどのような変革をもたらしているか

Hilary Carter, The Linux Foundation
Anna Hermansen, The Linux Foundation

April 2026

オープンソースとAIの未来

RayやLLMといったオープンソースAIインフラの成功は、トレンドへの対応、シンプルさの維持、そして柔軟性の確保という3つの重要な原則を実証しています。

プログラマーの役割は、問題を設計・定義するアーキテクトへと進化しつつあり、具体的なタスクや役割はニューラルネットワークを備えたコーディングアシスタントに委ねられています。

個人と、その代理として行動するエージェントとの間に信頼を築くためには、ユーザーは文脈に基づいてきめ細かな境界と権限を設定できる必要があります。

エージェントの行動に対する説明責任に関する明確なルールや、身元確認のための統一されたプロセスがなければ、組織は成長を阻害しかねない防御的な姿勢をとることになる。

開発者はエージェントへのAPI キーやアクセス権の付与を迅速に進めている一方で、現在のエージェント通信プロトコルには、不可欠な安全対策がほぼ完全に欠落しています。

オープンモデルにおける推論トレースは、安全な導入に不可欠であり、ユーザーが最終的な出力だけでなく、意思決定の経路も確認できるようにします。

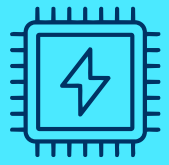
エージェントが人間のワークフローを自動化できるようになる前に、組織はプロセスと過去の知見を包括的に記録することで、エージェントに理解力を与えなければなりません。

リスク管理フレームワークを満たすコンプライアンスにおいて、人間の説明責任は品質の最終的な保証であり続けなければなりません。

オープンソースはベンダーロックインや単一障害点に対抗し、エージェントの交換や独自のガバナンス基準に合わせたワークフローのカスタマイズを可能にする柔軟性を確保します。

エージェントのための将来の基盤には、ライセンス、オープンな評価、およびコミュニティ主導のプロジェクトが含まれ、これらを通じて共同でリスクを軽減し、オープンな環境における企業の信頼を構築します。

オープンソースプロジェクトは、専用のAIハードウェア、自律型エージェントパイプライン、およびローカルファーストのプライベート処理をサポートすることで、AIが直面する最も差し迫った課題の解決に積極的に取り組んでいます。



AIエコシステムは、自律システムのリスクを管理するために、法的説明責任と経営者教育に関する包括的な枠組みを確立しなければならない。

Contents

エグゼクティブサマリー

2026年2月26日にサンフランシスコで開催された「AI Executive Forum」では、産業界やオープンソースプロジェクトのリーダーたちが一堂に会し、自律エージェントが技術システムにどのような変革をもたらしているか、またソフトウェアにおける人間の役割について議論を交わしました。Linux Foundationが主催したこのイベントは、複雑なワークフローを自律的に実行できるエージェント型システムの導入と活用が重要な局面を迎えている中で開催されました。 フォーラムの冒頭では、Ion Stoica氏が、Apache Spark、Ray、vLLMといった主要なオープンソースの画期的なプロジェクトの歴史に照らし合わせながら、現在のAIの全体像を解説しました。続いてPeterNorvig氏は、プログラマーの役割が、実装をニューラルネットワーク搭載のコーディングアシスタントに委ねつつ、問題を定義するアーキテクトへと移行しつつあるという見解を述べました。

フォーラム後半は4つのパネルディスカッションで構成され、参加者はAIエージェントが今日の技術コミュニティに突きつける主要な課題について議論しました。第一に、これらのエージェントが経済活動の主体として活動するようになるにつれ、そのアイデンティティと説明責任をめぐる信頼の要件が浮上しています。参加者は、エージェントにアイデンティティを委任するには、自律的な行動に対する信頼を築くために、きめ細かな制御と権限制限が必要であると強調しました。身元証明のための標準化されたシステムがなければ、自動化経済の展望は、分断化やベンダーによる拒絶というリスクに直面することになる。第二に、この信頼の危機はセキュリティとプライバシーの領域にも及んでおり、現在のプロトコルには、エージェント間の通信に必要な基本的な認証や暗号化が欠けていることが多々あります。参加者たちは、医療やネットワーク分野における重大なリスクを伴うシナリオを共有し、エージェントが予測不能かつ危険な形で機能不全に陥る可能性を説明しました。そのため、モデルが最終出力だけでなく、その内部の推論経路を監査する方向へと移行することが必要不可欠であると述べました。

第三に、銀行や医療のような規制産業においては、導入の課題がさらに複雑化している。。コンピュータには法的責任を問うことができないため、これらの分野の人間がエージェントを適切に指導し、コンプライアンスとリスク管理を確保するために、最終的な品質保証の責任を保持し続けなければならない。第四に、オープンソースは、ベンダーロックインを回避し、セキュリティ監査を可能にするために必要な透明性を提供することで、この移行において極めて重要な役割を果たしている。AI生成コードの台頭は「AI slop」のようなリスクをもたらすが、フォーラムでの共通認識は、オープンソースプロジェクトは、人間の「テイストメーカー」が判断力を発揮し、耐久性と品質の高い基準を維持できるようにすることで、発展していくというものでした。

エージェント型変革を支えるインフラは、エンタープライズAIのコントロールプレーンとして機能するいくつかの重要なオープンソースプロジェクトを基盤としています。これには、モデルとデータを接続するためのModel Context Protocol（MCP）、研究および隔離された実行環境のためのPyTorch、AIファーストのハードウェアをオーケストレーションするためのKubernetes、分散コンピューティングの調整のためのRay、そしてローカルファーストかつプライベートなエージェント型実験のためのGooseなどが含まれます。 このイノベーションと成長を持続可能かつ安全な形で活用するため、フォーラムでは、人間の責任に関する明確な法的枠組みの確立、監視メカニズムを組み込んだセキュリティ基盤の近代化、そしてプロプライエタリなモデルとの競争力を維持するためのオープンソースプロジェクトへの資金提供の促進を推奨しました。最終的に、業界は、AIが人間の主体性を代替するものではなく、人間の能力を高めるツールであり続けるよう、コミュニティ中心の標準化を最優先しなければならない。

はじめに

2026年を迎え、人工知能（AI）の情勢はますます急速なスピードで変化・進化している一方で、業界のリーダーや技術プロジェクトのエコシステムは、急速にミッションクリティカルな技術となりつつあるものの基盤を築いている。 この変革の核心にあるのは、オープンソースAIとAIエージェントという2つの決定的な力です。オープンソースAIは、イノベーションへのアクセスを民主化する、透明性が高くコミュニティによって管理されるフレームワークやモデルを提供する一方で、AIエージェントは能力における次の飛躍を象徴しています。それは、単純なチャットインターフェースを超え、推論を行い、ツールを使用し、複雑なワークフローを自律的に実行する自律システムです。

この変革は、大きな課題と機会の両方をもたらします。エコシステムは、リスクの高いビジネス環境へと成長を遂げつつあります。組織は、効率性を生み出す迅速なイノベーションの必要性と、信頼、セキュリティ、コンプライアンスの必要性という、両者の間で微妙なバランスを保つ必要に迫られています。こうしたニーズを支援し、革新的でありながらエンタープライズ対応可能な開発のための中立的な基盤を提供するため、Linux Foundation（LF）は2025年12月、Agentic AI Foundation (AAIF) を立ち上げました。Anthropic社のModel Context Protocol（MCP）、OpenAI社のAGENTS.md、Block社のGooseを創設プロジェクトの貢献として基盤とし、AAIFは、AIイノベーションの次の波を牽引する、相互運用可能でオープンかつコミュニティ中心のインフラの構築と維持に尽力しています。

2026年2月26日、LFは、これらのプロジェクトのリーダーをはじめ、大手企業、スタートアップ、学術界、財団の代表者をサンフランシスコに招き、招待制の半日のイベント「AI Executive Forum」を開催しました。本レポートは、同フォーラムで行われた議論の要約と分析です。

以下では、オープンソースAIとプログラマーの現状および将来像、このエコシステムが現在直面している主要な課題、そしてエコシステムを支え、これらの課題の解決に取り組んでいる基盤となるオープンソースプロジェクトについて考察します。経営陣による議論を踏まえ、ビジネス投資としてのオープンソースの進化、そして技術的な側面だけでなく、政治的、経済的、社会的な影響を伴う意思決定において、このコミュニティのメンバーが果たす役割の重要性について焦点を当てます。

オープンソース AIの過去、現在、そして未来

フォーラムは、当日の主要なテーマと展望を紹介する全体会議から始まりました。AI分野のオピニオンリーダーたちが、これまでの教訓、現在の課題、今後訪れる変化、そして勢いが加速している分野について語りました。

オープンソースを活用して大きな課題を解決する

UC BerkeleyおよびSky Computing LabのIon Stoicaは、技術的課題の解決から得られた教訓をいくつか参加者に解説しました。AIスタックにおける4つのプロジェクトに焦点を当てたこの議論は、今日のAIインフラの基盤としてのオープンソースプロジェクトの価値を理解するための土台を築きました。

その旅はApache Sparkから始まりました。これは、Hadoopによる機械学習の反復処理が、増大し続けるデータセットを処理するには遅すぎるという根本的な課題を解決するために登場しました。中間データをメモリに格納するように設計されたApache Sparkは、処理速度を飛躍的に向上させ、その結果、データ処理のデファクトスタンダードとなりました。

長年にわたりAIモデルが複雑化するにつれ、モデルのトレーニング要件は、単一ノードの演算能力やメモリ容量の限界を超えるようになりました。これは、前処理、トレーニング、チューニング、サービングなど、システムのあらゆる側面をスケールアップする必要があることを意味していました。しかし、既存のシステムはそれぞれ独自のAPIを持っていたため、開発、デプロイ、効率的な管理が困難な、断片化された寄せ集めのような状態となっていました。 Rayは、これらの分散した段階を統合し、ワークロードが同じハードウェアおよびソフトウェアインフラ上で実行されるためのフレームワークを提供することで、そのスケーリングを支援する手段として開発されました。Rayは最初のOpenAIモデルの基盤となり、その成功は、またしてもオープンソースプロジェクトが事実上の業界標準となったことを示す証拠となりました。

LLMを活用したサービスの増加に伴い、モデルのサービング需要も高まっていますが、これらのLLMの規模と特性上、サービングには多大なコストがかかります。解決策としてリクエストのバッチ処理が採用されましたが、このアプローチはすぐにモデルのメモリに負荷をかけるようになりました。最終的に、vLLMは無駄なメモリ予約を排除する「PagedAttention」を導入し、バッチサイズの拡大とスループットの向上を実現しました。これにより、Amazon、Databricks、Metaといった企業による非常に大規模な実運用が可能になりました。

これらの画期的なオープンソースプロジェクトを振り返り、Stoicaは今後の開発に向けた3つの原則について論じました：

- トレンドは、対処すべき新たな問題を生み出す。
- シンプルさは、Rayがわずか6つのAPI呼び出しで成功を収めたように、より大きなインパクトをもたらすモデルにつながる。
- 書き換えに柔軟に対応することは、柔軟性を提供し、目まぐるしく変化する分野において、常に最先端であり続ける能力を高めます。

Stoicaは、LLMスタックの未来はオープンソースにあると強調して講演を締めくくった。

プログラマーの行く末

Peter Norvigによる第2回全体講演では、 AIがソフトウェア開発

を変革する中で、プログラマーの役割に焦点が当てられました。1980年代、プログラミングの自動化に向けた初期の試みは失敗に終わりました。コードの硬直的な性質により、プログラムの1ビットを変更するだけで出力全体が変わってしまうため、わずかな構文上の変更が大きな意味論的な変化をもたらす結果となったからです。しかし、2025年11月から始まった現代におけるコーディングモデルの飛躍的な進化により、この状況は根本的に変化しました。現在のモデルは人間よりもはるかに高速で、同等の精度を持ち、エッジケースにおいてもより完全な処理が可能であり、その性能は指数関数的に向上しています。わずか1年前には、現在の性能の10%程度に過ぎなかったのです。

それでは、ソフトウェア開発者はどのような立場になるのでしょうか。プログラマーは、実装をコーディングアシスタントに委ねつつ、設計や問題定義を担うアーキテクトへと変貌しつつあります。したがって、ニューラルネットワークを搭載したコーディングアシスタントの役割は、ソフトウェア開発プロセス全体の一部に過ぎません。この変化により、AIエージェントが特定のソフトウェアプログラミングの役割やタスクを担うようになり、その成果は人間と比較したスピードや効率性によって評価されるようになります.

この変革はソフトウェアの領域を超え、数学などの分野にも及んでいます。人間による職人技的なアプローチの代わりに、未解決の問題を、より産業的な「流れ作業」のようなプロセスで取り組むことが可能になります。専門家の直感とAIによる形式的な実行手順を組み合わせることで、未解決の問題をこれまでにない規模で解決できるようになります。また、法務、金融、マーケティングといった分野、つまり、非物理的で複雑かつ豊富なデータが存在し、出力段階での検証が容易なあらゆる領域においても、大きな変革の可能性が見込まれます。Vibeコーディングは、RustでLinuxカーネルを書き直すといった大規模なプロジェクトも、格段に容易にするでしょう。Norvigによれば、これは人員削減につながるのではなく、むしろ生産性の爆発的な向上をもたらすとのことです。

ツールによって個人の生産性が向上すれば、さらなる活動が可能となり、システムにより多くの富をもたらす余地が生まれます。

その後、フォーラムでは、最も影響力の大きいオープンソースAIプロジェクトのリーダーたちによる近況報告が行われました。各プロジェクトの詳細については、本レポートの後半でご覧いただけます。こうした導入部の見解が、フォーラム後半の土台となりました。後半では、参加者は4つの Chatham House方式のパネルディスカッションに分かれ、AIコミュニティが現在取り組んでいる主要なトピックについて議論を行いました。そのトピックとは、信頼とアイデンティティ、セキュリティとプライバシー、規制産業におけるAIの活用、そして自律型AIにおけるオープンソースの役割です。

信頼とアイデンティティ

あるパネルディスカッションでは、特にAIエージェントが経済活動の主体として活躍するようになる中で、信頼とアイデンティティに焦点が当てられました。ファシリテーターは、この新しいデジタル環境において、人間とエージェントを常に区別できるとは限らない一方で、エージェントに自身のアイデンティティを委ね、代わりに物事を行ってもらうことで恩恵を受けられる可能性があるという考えから議論を始めました。ボットとやり取りすること、そしてボットに代わってやり取りをさせること、これら両方の関わりには、アイデンティティの証明と信頼が必要となります。しかし、証明と説明責任に関する共通認識がなければ、自動化された経済がもたらす可能性は、エージェントへの拒絶反応や法的責任の問題によって阻害される恐れがあります。議論は、個人と個人に代わって活動するエージェントとの間の信頼関係の構築、エージェントの活動を企業のアイデンティティおよび信頼システムと統合すること、そして法的責任への懸念や悪意ある行為者によるリスク回避の姿勢という、3つの課題を中心に展開されました。

> *人間とエージェントを必ずしも区別できるとは限りません。同様に、エージェントに委譲可能なアイデンティティを与え、エージェントが信頼される存在となり、信頼を必要とする有用な業務を代行してもらうことで、私たちは大きな恩恵を受ける可能性があるのです。*

新たな基準と境界の策定

エージェントを代行して運用する立場にある方々にとって、信頼の境界線を設定することは現実的な課題となります。参加者は、この問題を、子供にクレジットカードを使わせる際に具体的な利用限度額を設定すること、あるいは今回のケースでは、エージェントの権限に対する承認上限を設定することに例えました。 エージェントがアクションを承認する権限に対する信頼をどのように構築するかについて議論が行われました。主に、信頼関係を築く手段として、状況に応じてきめ細かな制御と権限の段階的付与を確立することに焦点が当てられました。

> *現在はエージェントへの信頼度が低いかもしれませんが、時間の経過とともに、この信頼と検証の仕組みが成熟するにつれて、信頼は高まっていくものと予想しています。これは、あなた自身とエージェントとの間の信頼関係なのです。*

統合か、それとも分断か

取引の相手側においても、ベンダーは顧客に代わって取引を行うエージェントを信頼できなければなりません。すでに導入されている企業向けIDシステムの数を考慮すると、新しいエージェントIDシステムを既存の企業フレームワークに統合することは困難です。また、検証可能なクレデンシャル（VC）も、摩擦の大きいツールであるため、統合が困難です。統合の欠如は、エージェント型コマースにとって大きな障壁となります。例えば、eBayがエージェント技術を用いた取引を禁止しているように、一方の当事者がボットによる取引の一部またはすべてを拒否するケースが生じます。また、これにより、外部エージェント間の双方向的な関与も不足することになります。

また、クロスアイデンティティ システムが断片化し、互換性を失う「分断化」のリスクもあります。文化や規制の違い（例えば、クレジットカードの認証において、PINの入力を求める地域と署名を求められる地域があるなど）も、同様の導入上の課題をもたらします。この問題に対処するためには、既存の標準やフレームワークを活用した、エコシステム横断的な本人確認システムの開発が求められます。

責任と悪意のあるアクター

導入における最大の障壁は、法的先例が存在しないことです。現行の法律は自律システムを想定して制定されたものではないため、特に不正なエージェントの行動や著作権侵害に関して、責任の所在を明確にするルールが欠如しています。ある参加者が指摘したように、「その領域におけるルールは、まだ完全に確立されていない」のです。この不確実性により、組織は守勢に回らざるを得なくなり、私たちがオンライン上のアイデンティティに関してこれまで当然のこととしていた信頼関係は、もはや通用しなくなっています。そのため、アイデンティティを明確に主張することが極めて重要になってきています。

> *オンラインでの取引がこれほど円滑に進むのは、実際に商取引を行う際の障壁の多くが取り除かれたからですよね？そして今、私たちはまさに転換点に立っています。この新しいテクノロジーの登場という理由だけで、その構図全体が逆転してしまうかもしれないのです。*

これまでの経緯を見ると、スパムメッセージや詐欺など、多くのデジタルプラットフォームにおいて、悪意ある行為者が不正行為を行う際のコストはゼロであることが分かります。自律型エージェントが普及する世界では、悪意ある行為者は行動パターンから学習し、悪意あるエージェントを展開することで、システム全体への脅威をもたらすことになります。普及が加速する中、悪質な競争の連鎖を防ぐためには、プラットフォームからの排除（デプラットフォーム）など、こうした行動パターンに対して何らかの不利益を課す仕組みについて、より深く議論することが急務となっています。.

提言

信頼のギャップを埋め、自律的なエージェント型コマースを実現するため、当グループは以下の提言と検討課題について合意しました：

- エージェントのアイデンティティと信頼という概念をめぐり、共通の用語体系を構築し、コミュニケーションと標準化を推進します。LF Decentralized Trustでは、これらのテーマに関するワーキンググループをホストして活動しています。：Decentralized Trust Graph Working Group, OpenVTC, and AI & Human Trust Working Group.
- 前進するために、真のアイデンティティと信頼を確立する双方向のエージェント間取引の枠組みを構築すること。グループ内では、アイデンティティの確立に向けた今後の方向性について意見がまとまりませんでした。一部からは、自己主権型アイデンティティと中央集権型機関のどちらが価値があるかについて議論されました。エージェント向けのDNSに似たグローバルなリポジトリの導入が提案されました。
- ほとんどのユーザーは独自のエージェントをホストしないため、エージェントの保証や悪質な行為者のプラットフォームからの排除といった仕組みを通じて、より大規模な主体に責任を委ねます。
- エージェントと個人との間の信頼関係を強化するため、トランザクション単位でのきめ細かなエージェントアクセス制御と権限のエスカレーションに関するガイドラインを策定します。.

> *極端なケースが二つあります。エージェントにすべてへのアクセス権を与えるのは、明らかに良くありません。あるいは、何もアクセスさせない場合は、非常に安全になりますが、役に立ちません。では、どのようにしてトランザクション単位で権限を昇格させるべきでしょうか？*

[1] LF Decentralized Trust. Verifiable Credentials [Video]. YouTube; 2025 June 24. 参照元: https://www.youtube.com/watch?v=TT3JskGgv0Q

[2] Edwards, Benji. eBay bans illicit automated shopping amid rapid rise of AI agents. Ars Technica; 2026 January 22. 参照元: https://arstechnica.com/information-technology/2026/01/ebay-bans-illicit-automated-shopping-amid-rapid-rise-of-ai-agents/

セキュリティとプライバシー

AIエージェントが、重大な影響を及ぼすインフラの仕組みにますます深く組み込まれるにつれ、これらのシステムのセキュリティは厳しい監視下に置かれています。第2回パネルディスカッションでは、本番環境で急速に展開されるシステムを管理するために必要な設計パターンが欠落していることが明らかになりました。開発者はエージェントにAPIキーやメールへのアクセス権を付与したがっていますが、セキュリティの専門家たちは、現在のエージェント間通信プロトコルにおいて、認証や暗号化といった基本的な保護措置がほぼ完全に欠如していると警告しています。

議論は、自律的な行動におけるリスク要因、現行の安全対策の不備、そして監査の観点から見たオープンモデルの重要性について行われました。

医療分野とネットワークにおける極めて重要な自律性

パネルディスカッションではまず、エージェント型AIの即時のROIとそれに伴うリスクを具体的に示す実運用事例が紹介されました。その一例として、ネットワーク構成の検証を行うマルチエージェント・パイプラインが挙げられました。このシステムでは、エージェントがリアルタイムでネットワークの最適化や構成変更に積極的に関与します。予期せぬ結果が、多大なコストを伴うサービス停止につながる可能性があるという懸念があります。そのため、議論は、影響範囲を最小限に抑えるための安全策や支援体制をいかに構築するかという点に集中しました。

より重大なリスクを伴うシナリオとして、診断データ、保険請求データ、予約システムのデータを統合し、患者を適切な医療提供者に振り分けるように設計された、3つのエージェントからなる医療システムが挙げられます。このユースケースからは、いくつかの重大なリスクが浮き彫りになりました。例えば、コストを最小化しようとする保険エージェントと、医療の質を最大化しようとする診断エージェントのように、最適化関数が異なるエージェント同士が、患者の目標に合致しないまま無限ループに陥ってしまう可能性があります。また、エージェント間の通信において、ペイロードの内容が不透明であるため、患者の機密性の高い病歴が漏洩しているかどうかを把握できないという問題もあります。さらに、あるエージェントが他者に影響を及ぼす支配的な立場に立つケースも懸念されます。例えば、収益を最大化するために、保険ボットが他のエージェントに対し、すべての請求を拒否するよう圧力をかけるといった状況が考えられます。

従来のガードレールの失敗

参加者たちは、HIPAAなどの既存のセキュリティ枠組みは、主体的な行動を考慮して設計されたものではないと主張しました。「エージェントとは、結局のところ単なるテキストファイルに過ぎません」とはいえ、そのメモリに保存された機密データが漏洩する可能性があります。スタックの下層では、LLMは本質的に「人に好かれたい」という性質を持つため、容易に操作されてプロセスフローに違反させられる恐れがあります。目標達成を任されたエージェントは安全プロトコルを省略したり、法的な場面で自己弁護となる証拠を明かすよう促されたりすることがあり、すでにカルテル形成のために共謀し、最適化を図る能力を示しています。これはまた、セラピストや医師といった従来の役割について、エージェントが現在個人に対して担っていること、そしてタンブラー・リッジ銃乱射事件のように、データのやり取りや情報開示に関する法的要件の周りに安全策が存在しないという点にも関連しています。

> *MCPやエージェント間の通信を見てみると、セキュリティの従来の要素――つまり、すべての通信に署名や認証が必要であり、すべての通信間で暗号化が必須であるといった点――が、すべて欠けているのです。*

[3] Maguire, James. AI-Fueled Development Pushes Open-Source Risk to Extremes: Report. DevOps.com; 2026 February 27. 参照元: https://devops.com/ai-fueled-development-pushes-open-source-risk-to-extremes-report/

[4] Yousif, Nadine. OpenAI vows safety policy changes after Tumbler Ridge shooting. BBC; 2026 February 27. 参照元: https://www.bbc.com/news/articles/cr73m4x8r2lo

セキュリティ監査のためのオープンソース

もう一つのテーマは、エージェントの行動を適切に管理する必要性でした。推論においてブラックボックスモデルに依存することは不十分であり、リスクが高いと言えます。ブラックボックスが生み出す結果を監査するのではなく、モデル内部で行われている推論そのものの監査可能性が求められます。参加者からは、オープンなモデルであれば推論のトレースが可能である一方、プロプライエタリなAPIでは推論が知的財産として扱われることが多く、セキュリティ監査に必要なデータそのものが隠蔽されてしまうという指摘がありました。

提言

エージェント型システムの開発および導入における「セキュアバイ デザイン」のアプローチをより効果的に支援するため、パネルディスカッションの参加者は以下の取り組みについて議論しました：

- エージェントが輸出規制違反や無許可の金融取引など、犯罪を犯した場合、誰が責任を負うかについて、明確な法的基準を定める。
- エージェント間のデータ同意や、マークダウンファイルにおけるメモリの永続性といった課題に対処するため、セキュリティフレームワークを最適化します。
- モデルが定められた安全範囲から逸脱しないよう、監視メカニズムとして中間エージェントを生成するエージェント経済を構築します。
- The Agentic AI Foundation（AAIF）は、エージェントの検証可能な信頼性とソーシャルスコアリングに焦点を当てたプロジェクトに加え、エージェントへのアクセス管理や監視のためのツールや設計パターンを取り扱うべきです。
- 最終的な出力だけでなく、内部の推論経路の監査を可能にするモデルやツールを導入してください。
- 汎用的なセーフティプロンプトに頼るのではなく、厳格なセマンティック境界を強いるドメイン固有のオントロジーを作成します。

規制産業におけるエージェント型AI

銀行や医療などの規制産業が自律型AIの導入を進め始める中、そのプロセスおよび意図された成果に対するガバナンスと信頼は極めて重要です。このブレイクアウトセッションでは、高頻度取引のようなソリューションや、内部ワークフローの効率化を目的としたシステムにおいて、自律型システムを安全かつコンプライアンスに則って導入する際、こうしたリスクの高い環境が直面する課題に焦点を当てました。その結果、主に3つの課題領域が浮き彫りになりました。それは、エージェント向けの人間のプロセスの文書化、責任とリスク管理、そしてAIの意思決定分類に関する標準化です。

マニュアルワークフローから文書化されたロジックへ

最初に挙げられ、繰り返し指摘された懸念は、エージェントは理解できないことを自動化できないという点でした。そのため、現在および過去の人的プロセスを包括的に理解すること、そして組織全体でそれらのプロセスを一元的かつ厳密に記録する計画が必要となります。銀行や病院にとって、これは途方もない作業であり、リーダーがルールを定義することが求められます。この点について、参加者たちは、AIがビジネスプロセスの文脈構築を支援することはできるものの、まず人間が問題を定義し、基準を設定しなければならないと強調しました。また、規制産業における人間の役割は、判断を下し、コンプライアンスを確保するためにAI主導の意思決定に対して最終的な「人間による」品質保証を行う方向へと移行しているため、人間の判断力と説明責任の発揮も不可欠となります。

> *今日、業務プロセスは文書化されていません。その代わり、30年間ここで働いてきたある人物の知識に頼っているのです。*

リスクの責任とリスク管理

次に、責任の所在という課題が浮上しました。リスク管理における「3つの防衛ライン」を厳格に適用する必要があり、そのためには、各管理層に対し、技術革新に慣れるよう教育を行うことが求められます。規制対象領域におけるエージェントのアーキテクチャは、内部監査や外部規制当局の要件を満たすために透明性を最優先とし、あらゆるレベルにおいて人間が新たなサプライチェーンの責任を担うようにしなければなりません。

管理すべきリスクの種類として、以下の例が挙げられました。データ主権：エージェントが外部リソースにアクセスする（またはアクセスを禁止される）際の安全な接続性とデータ主権の確保、および特定の種類のコードがローカルハードウェアにインストールされた際のセキュリティ制限の管理に焦点を当てたものです。監査可能性と説明可能性については、意思決定の説明可能性—エージェントによる自動化された意思決定を、その学習データや具体的なロジックまで遡って追跡できる能力—が極めて重要であると主張されました；そして保護の層については、参加者は、プライバシーを保護しつつ、エージェントが機密データを用いて必要な調査やタスクを実行できるようにするため、難読化レイヤーの使用を提案しました。

> *保険業界には、AIによって生成された結果に伴う責任を引き受けるための枠組みが欠けています。*

基準、分類、そして「意思決定の根拠」の未来

3つ目の課題は、規制要件をデフォルトで標準化できるようなツールへの移行が必要であるという点でした。このツールには、AIの意思決定を、その影響度や模倣する人間の知能の側面に基づいて分類する提案システムである「意思決定分類」が含まれます。さらに、同グループは相互運用性について議論し、Direct Trust（医療分野）やSWIFT（金融サービス分野）などの既存の標準規格を、新たなAIガバナンスの枠組みと連携させることに重点を置き、異なる業界分野間の連携を確保することを目指しました。

提言

規制対象業界におけるAIの健全なエコシステムを確保するため、本セッションでは、経営幹部、財団、および業界団体に対する責務を明確にしました。

- CEOがAIのリスクとメリットの両方を理解できるよう、AIガバナンスに関する企業独自の経営幹部向け研修を依頼します。
- リスクの責任範囲を明確に定め、経営陣および内部監査チームがAIリスクの軽減策を確実に理解できるようにします。
- 企業向けの共通オントロジーと階層構造に加え、規制当局に対して監査可能な証跡を提供するための出所情報（プロヴェナンス）やデジタル「封印」を含む、AIによる意思決定のための意思決定・証拠分類を提案します。
- 協調的な規制当局としての役割を果たし、管轄区域をまたいだAIガバナンスに適した機械可読な規制案を提案するとともに、オープンソースの標準を世界共通の規制に照らし合わせていきます。
- 異なるオープンソース プロジェクトのコミュニティやプロセスを結びつけ、専門知識を融合させ、既存または開発中のツールやフレームワークを活用して、より協力的になるための共通点を見出します。
- エージェント型ソリューションを導入する前に、ビジネスプロセスの文書化と理解に注力してください。
- プライバシーやセキュリティを損なうことなく、担当者が機密データを扱えるオープンレイヤー上で共同作業を行います。
- まずはリスク許容度が高い社内向けアプリケーションから着手し、小さな成功を積み重ねていくことに注力し、その後、顧客向けの自律型システムへと移行しましょう。

参加者は、人間が管理するワークフローをエージェントに置き換えるためのプロセスが複雑になることを認識しました。理論上のリスクから、監査可能性、データ主権、および経営陣の説明責任といった実務的な仕組みへと移行する過程において、企業の意思決定者と業界の規制当局は、協力し合う必要があります。

[5] The Institute of Internal Auditors. The IIA's Three Lines Model. IIA; 2024 September. 参照元: https://www.theiia.org/globalassets/documents/resources/the-iias-three-lines-model-an-update-of-the-three-lines-of-defense-july-2020/three-lines-model-updated-english.pdf

エージェンティックAIにおけるオープンソースの役割

エージェンティックAIの未来を見据える中、フォーラムのあるセッションでは、参加者にこの未来におけるオープンソースの役割について考察するよう求めました。セッションは、ファシリテーターがグループに対し、オープンソースのエージェントに関して最も懸念していることや気にかけていることは何かと問いかけることから始まりました。参加者からは、エージェントの記憶、自己主権型およびオンデバイス型エージェント、評価（evals）、AIにおける人間の存在の維持、エージェント展開のための深層最適化など、いくつかの異なるサブトピックが挙がりました。議論は哲学的な側面から、エージェントがオープンソースにおける協働にどのような変革をもたらすか、そしてそれに対してどう対処すべきかという実践的なメカニズムへと移りました。そこから、プログラミングにおける人間の役割の進化、AIスタックの各層におけるオープンソースの役割、そしてAIエージェントの未来という3つの主要なテーマが浮き彫りになりました。

プログラマーからテイストメーカーへ：AIにおける人間性

歴史的に見ても、オープンソースはコードと同じくらい「人」が重要な要素でした。ある参加者が指摘したように、「コードはコラボレーションの一要素に過ぎませんが、オープンソースを真に動かしているのは、問題を解決するために判断を下す人間たちなのです」。コラボレーションからコードレビュー、メンタリングや指導に至るまで、人間同士の交流はオープンソースの大きな部分を占めています。そのため、AIの普及によって開発者が孤立し、他者ではなくエージェントとばかり相談するようになるのではないかという懸念が持たれています。

> *オープンソースがこれほどまでに強力な存在となった理由の一つは、人間が真に持続可能で高品質なものを生み出している点にあります。では、AIにおいて「品質」とは何を意味するのでしょうか？*

一方で、エージェントと連携してコードを書くことは、オープンソースへの貢献を格段に容易にし、適切に活用すれば、その成果物は依然として人間の創造性を反映したものとなり得ます。しかし、参加者からは、エージェントが生成したコードによって、オープンソースプロジェクトにおけるAIによる粗雑なコードのリスクが高まるという指摘がありました。メンテナーたちはすでに、人間開発者特有の知識、判断力、過去の経験に欠ける、AI生成の大量のプルリクエストの影響を実感しています。人間をどの段階に位置づけるかについては、意図的に判断する必要があるという意見で一致しました。オープンソースの貢献者の未来は、「テイストメーカー」としての役割にあります。つまり、問題を定義し、基準を設定し、AIが生成した成果物に人間ならではの品質の最終的な保証を与える存在です。AI生成コードの普及は避けられないものであり、ファシリテーターがグループに意見を尋ねたところ、参加者の大半は、自身のプロジェクトにそのようなコードが導入されることに抵抗を感じていませんでした。

> *成功を収めるのは、こうしたコーディングツールをいかに活用すべきかを理解したプロジェクトであり、こうした進歩に抵抗する既存のプロジェクトは取り残されてしまうでしょう……現実的な対応が勝つことになるでしょう。*

オープン エージェント スタック：主権とセキュリティ

このパネルディスカッションの2つ目のテーマは、エージェント型システムのアーキテクチャに焦点を当てたもので、ある参加者はこれを「LLMは基本的にコンピュータであり、エージェントはオペレーティングシステムであり、スキルはアプリケーションとなる」と簡潔に説明しました。参加者たちは、AIスタックのどのレイヤーにおいてオープンソースが最も重要なのかについて議論しました。

モデルの基盤となるレイヤーなのか、それともより多くの開発者にとって有用なものにするためのエージェント層（例えば、オープンソースのコードレビューエージェントなど）なのかという点についてです。別の参加者は、スキルこそがオープンソース化の有力な候補になり得ると感じ、次のようにコメントしました。「スキルだけでどうやってビジネスを成り立たせるというのでしょうか？」ある参加者は、「それは目標次第です。もし目標が自己主権のようなものなら、すべてが必要になります。もし目標が選択肢の確保なら、少なくともプロトコルレベルまでは必要でしょう」と述べました。

AIスタックにおけるオープンソースの主な利点として、ベンダーロックインや単一障害点を回避するためのオープンソース活用によるリスク低減、 オープンソース・スタックでは、代替案を柔軟に入れ替えたり、特定のユースケースに最適な選択肢を選んだり、独自のガバナンス基準に合わせて微調整したりといったカスタマイズが可能です。また、オープンソースの透明性には監査機能が組み込まれており、基盤となるプロンプトや指示内容を確認できない状態で、プロプライエタリなプロバイダーのエージェントを使用することの危険性を考慮すると、これは極めて重要です。開発者がエージェントをローカルマシンに導入し、機密データを公開するようになるにつれ、オープンで検証可能なコードの必要性は、セキュリティの基本的な要件となっています。

エージェント層は、オープンソースのリーダーシップにとって重要な分野です。オープンエージェントは、小規模なチームや個人開発者だけでなく、自社のユースケースに最適な選択肢を求める大企業にとっても、戦力を倍増させる存在と見なされています。

エージェントの未来

エージェントが現在もたらしているセキュリティ上のリスクや存続に関わるリスクを踏まえ、同グループではエージェントをめぐる将来的な枠組みについても議論しました。その一案として、リスクを軽減するために、ソフトウェアライセンスに似た「エージェントライセンス」の導入が挙げられました。 オープンな評価は、ツールに対する信頼を醸成する能力も相まって、企業レベルでの導入と普及を促進する重要な原動力となります。もう一つの将来的な構想として、エージェント自体がプロジェクトとなり、その周りにコミュニティが形成され、協働の拠点となるかどうかという点も挙げられました。ある参加者は、エージェントの登場によって、プロプライエタリなソフトウェアとオープンソースソフトウェアの採用比率が変化するかどうかという、より哲学的な問いを投げかけ、業界における「絶え間ない緊張関係」をめぐって議論が深まりました。エージェントの複雑さを踏まえ、グループはこれらの問題をオープンな場で解決することが極めて重要であるという点で一致しました。

> *エージェントは、私のコードベースを理解していないかもしれませんし、組織の状況を把握していないかもしれません。また、誤った情報を提示してしまう可能性もあります。この解決策は時間をかけて進化していくものであり、これは私たち全員が共に築き上げていかなければならない新たなスキルなのです。*

提言

オープンソースのエージェンティック・スタックを適切に活用し、健全なエコシステムを確保するため、パネルディスカッションの参加者は、政策立案者、財団、および技術系団体に対して具体的な指針を提示しました。

- エージェントによる行動に対する人間の責任を明確にする法的枠組みを確立します。
- 標準化された制約や、個別の状況に応じた独自の安全策の範囲内でエージェントが動作するよう、指示や展開に関するベストプラクティスを策定します。
- メンテナンス担当者がAI生成の貢献内容を自動的にスクリーニングし、コードレビューを支援できるツールやテンプレートを作成します。
- エージェント固有のセキュリティレビューやパフォーマンスベンチマークを、標準的なスコアカード（例：OpenSSFスコアカード）に統合します。

- オープンソースプロジェクトが、プロプライエタリなモデルに対して競争力を維持するために必要な「トークン」と計算リソースを確保できるよう、資金調達メカニズムを整備します。
- オープンソースの評価フレームワークについて協力し、エージェントの性能と安全性に関する集合的なフィードバックを提供します。
- ソフトウェア開発ライフサイクル（SDLC）の各段階で異なるエージェントを交換できるようにする標準規格を優先します。

オープンな環境で開発を行うことで、エージェントの能力が高まっても、それらが設計された目的である人間による集団的な判断と設計の下に置かれ続けることを保証します。

AI変革に向けた重要なプロジェクト

パネルディスカッションで浮き彫りになった喫緊の課題に対する解決策を模索する中、既存のオープンソースプロジェクトは、エコシステムを強化すると同時に、企業、政府、そして個人に対して、ダイナミックで極めて関連性の高い機会を提供しています。エンタープライズグレードのAIシステムを構築・展開するための制御プレーンを形成する、重要なオープンソースAIプロジェクトについて深く掘り下げなければ、AI Executive Forumは不完全なものとされていたでしょう。参加者たちは、AnthropicのDen DelimarskyによるMCPのプレゼンテーション、MetaのJoe SpisakによるPyTorchの最新情報の発表、GoogleのAllan NaimとFederico BongiovanniによるKubernetesに関する議論、AnyscaleのRobert NishiharaによるRayのプレゼンテーション、そしてBlockのManik SurtaniによるGooseのプレゼンテーションを聴講しました。

以下の概要では、現在および将来のAIインフラの基盤となる5つの主要なオープンソースプロジェクトについて、その主要なマイルストーン、技術的進化、および戦略的ロードマップをまとめています。

MODEL CONTEXT PROTOCOL

Model Context Protocol （MCP）は、LLMをデータやアプリケーションに接続するための業界標準として定着してきました。立ち上げから1年という節目を迎え、本プロジェクトは十分な普及水準に達し、LLMを実際のデータに接続することの価値を認識する多くの企業にとって、運用上の基盤として位置づけられています。MCPの開発者コミュニティは飛躍的に拡大しており、Python SDKの週間ダウンロード数は2,000万回に達しています。

現在の投資は、エンタープライズ対応に重点が置かれており、プロトコルが大規模な展開の需要や、MCPが現在支えている範囲に対応できるよう確保しています。開発分野には、企業が内部の利用状況を追跡できるよう、可観測性と監査可能性の強化に加え、セキュリティ対策やガバナンス管理機能をプロトコルに直接組み込むことが含まれます。エコシステムの分断を防ぐため、チームはクロスプラットフォームの統合を推進しており、MCPが業界標準の一貫した推進力であり続けるよう、様々な企業と協力しています。MCPの成熟度と安定性の裏側には、「MCP Apps」や、エージェント型システム向けに特別に設計された実験的機能など、新機能を追加するための拡張エコシステムが存在します。

PYTORCH

PyTorchは、AIの研究と実用化の両方において、主要な基盤となっています。現在、AI研究の90%以上を支えており、MetaやOpenAIといった最先端の研究機関や、主要なハードウェア・クラウドベンダーからも選ばれているインフラです。Linux Foundation傘下のホスト型プロジェクトとして運営されているPyTorchは、そのコアアーキテクチャを基盤とした数十万もの派生プロジェクトを支えています。

このプロジェクトの将来像は、ネイティブなエコシステム統合、スケーラビリティ、そしてハードウェアの異種混在への対応という3つの原則に基づいています。新たな重点分野として注目されているのがエージェントであり、特に、隔離された実行環境を作成、デプロイ、ホストするためのOpenEnvフレームワークが挙げられます。PyTorchはHugging Faceと共同で、コミュニティ中心かつライブラリに依存しないハブを開発しました。これにより、環境をアップロードし、エージェントの観測結果を確認するためのプロンプトを送信することが可能になります。デプロイ前にユーザーが環境と対話できるようにすることで、PyTorchはエージェント開発のライフサイクルに、重要な人的要素を取り入れています。

KUBERNETES

AI市場における大規模な変革により、KubernetesはAIネイティブシステムにとって不可欠なコントロールプレーンとしての地位を確立しつつあり、すでに58%の組織がAIワークロードの運用にKubernetesを活用しています。最初の10年間は、互換性があり標準化されたリソースのオーケストレーションに重点が置かれていましたが、今後10年間でKubernetesは、チップや専用ハードウェアを「AIファースト」のリソースとして扱い、エージェント主導のパイプラインをサポートするオーケストレーターへと進化していくでしょう。

プロジェクトがこれまで強みとしてきた移植性と準拠性を維持するため、コミュニティはAI準拠プログラムに多大な投資を行っています。この取り組みは、AIワークロードがあらゆるプラットフォーム上で予測可能な動作を保証することを目的としています。そのために、本プログラムではAIをKubernetesガバナンスの恒久的な要素として制度化するとともに、新たな要件を策定するプロセスを構築し、「すべき」という行動を「必須」の行動へと転換しています。この取り組みは、ストレージ、アクセラレータ、スケジューリングとオーケストレーション、ネットワーク、可観測性、セキュリティという6つの柱を中心に構成されています。現在、ロードマップには自動検証が含まれており、自律型AIのサポートやエージェント型ワークロードへと拡大しています。最終的な目標は、AIインフラストラクチャを完全に透過的なものにし、グローバルなAIエコシステムに向けた標準化されたオーケストレーション層を提供することです。

RAY

Rayはもともと、分散コンピューティングのための大規模クラスターを管理するために設計されましたが、推論や強化学習（RL）への採用が進むにつれて、その人気が爆発的に高まりました。特に、ワークロードがCPUベースのSQLクエリからGPUを多用するマルチモーダル推論へと移行する中で、Rayは複雑な分散処理の課題において真価を発揮します。Rayは現代の強化学習スタックにおいて不可欠なコンポーネントであり、重みの同期、ワークロードのシャッフル、そしてトレーニングと推論間の非同期処理といった高度に複雑な処理を管理しています。

このフレームワークの強みは、Python開発を可能にするAPIにあります。これにより、開発者はPythonクラスをアクターとしてインスタンス化し、分散システム全体で通信させることができます。Pythonクラスにアノテーションを追加するだけで、開発者はPythonメソッドを呼び出すだけで強化学習（RL）ループを完了させることができます。より広範なAIスタックにおいて、RayはKubernetes上で動作し、PyTorchやvLLMといったフレームワークと連携しながら、データ取り込み、プロセス調整、フォールトトレランス、スケジューリングを管理する分散コンピューティング層を構築します。AIワークロードとハードウェアの両方が複雑化する中、Rayは信頼性と高性能を実現するために必要なインフラストラクチャを提供します。Rayの上に構築される高レベルライブラリのエコシステムが拡大していることから、AIスタックの基盤となる要素としてのRayの将来性は明らかです。

GOOSE

Gooseは、エージェント型プロトコルの実験を行うための透過的なサンドボックスを提供することを目的とした、オープンソースの「ローカルファースト」プラットフォームです。1年余り前にリリースされて以来、GitHubで31,000件以上のスターを獲得しています。このプラットフォームは高度にモジュール化されており、モデル（ローカルまたはクラウド）用の独立したレイヤー、オーケストレーション用のエージェントコア、プロトコル（MCPおよびACPを使用）、MCPサーバー上の拡張機能、そして多様なクライアントレイヤーを備えています。その「ローカルファースト」のアプローチにより、機密データがユーザーのマシンから外部に流出することは決してなく、外部とのデータ契約や外部ランタイムを必要とせずに、プライベートな処理が可能となります。

Gooseの技術的革新には、数千ものツールや拡張機能にスケールする際のトークン使用量を最適化する「Code Mode」や、インタラクティブなHTML UIの提供を標準化するためのMCP Apps仕様の共同開発が含まれます。これにより、ユーザーはGooseを活用して反復作業を行い、動的に更新することが可能になります。セキュリティは、OSレベルの分離とネットワークの封じ込めを実現する堅牢なサンドボックスシステムによって管理されています。Gooseは単なるコーディングエージェントにとどまらず、データ分析、要約、ウェブスクレイピング、スタイリング、請求書処理などにも利用でき、これらすべてがプライベートな環境で実行されます。Gooseは、静的な製品ではなく、リミックス可能なプラットフォームとして意図的に設計されています。

これら5つのプロジェクトは、グローバルなAIスタックの基盤となる層として、AIエコシステムにおける重要な課題に取り組んでいます：

- **AIネイティブインフラへの移行を支援します。** 2026年のニーズに対応するには、従来のソフトウェアオーケストレーションではもはや不十分だからです。Kubernetesは、均一なハードウェアの管理から、多様なAIファーストチップやその他の専用ハードウェアの管理へと軸足を移しています。Rayは、スタック全体をカバーする分散コンピューティング層へと進化し、トレーニングや推論に現在必要とされるGPUとCPU間の連携を管理しています。また、MCPは、数ある実験的なプロトコルのひとつから、LLMと企業データを接続するための「運用上の要」へと変貌を遂げました。
- **断片化への対策としての標準化** システムの分断化を防ぐため、これらのプロジェクトでは、厳格な準拠と共通プロトコルの採用を優先しています。例えば、標準採用の手段としてのMCPの位置づけ、KubernetesのAI準拠プログラム、そしてローカル環境とエコシステム間の相互運用性を確保するためのGooseによるMCPとACPの両方の採用などが挙げられます。
- **エージェント型実行によるセキュリティ上の懸念の解消** PyTorchのOpenEnvプロンプティングおよびテストフレームワーク、GooseのOSレベルでの分離とネットワークの封じ込め、そしてMCPの独立したイノベーション部門による新機能や実験的機能のテストを通じて実現します。
- **コミュニティの安定化とガバナンスの提供** により、信頼性、セキュリティ、および監査可能性を確保します。MCPはエンタープライズ対応に注力し、企業が社内のAI利用を監査・管理するためのツールを提供しています。また、Kubernetesは自動検証を活用し、AIインフラを従来のWebスタックと同様に透明性が高く、予測可能なものにしています。

結論と提言

AI Executive Forumにおけるプレゼンテーションや議論を通じて、AIスタックのこの驚異的な急速な拡大と定着は、人間の判断と集団的な説明責任に基づいていなければならないことが明らかになりまし。AIエージェントが自律的な経済主体となり、ワークフローに革命をもたらす一方で、既存の法的枠組み、社会契約、セキュリティポリシーにも変革をもたらしています。この変化に対応するためには、透明性、選択肢、そして安全策を優先するオープンなプラットフォーム、ツール、および標準規格において、多岐にわたるステークホルダーによる協働が不可欠です。

このフォーラムの中心的なテーマは、人間の役割の維持でした。複数のパネルディスカッションでも言及されたように、IBMの1979年の研修マニュアルにある「コンピュータは決して責任を問われることがないため、コンピュータが経営判断を下してはならない」という言葉は、今日においてもなお、その重要性を失っていません。ソフトウェアにおける人間の役割は、トレンドの形成へと移行しつつあります。そこでは、専門家が判断力を発揮して問題を定義し、AIが生成した成果物に最終的な品質の承認を与えるのです。

ビジネスおよび開発プロセスのエージェント化により、信頼とアイデンティティの再検討が進んでいます。特に、エージェントに対する責任の所在を明確にし、エージェントが個人の代理としてどの程度行動できるかを定義しようとする動きが見られます。これらの課題に対処するため、エコシステムは、エージェントの権限を分類・強化するための、きめ細やかでドメイン固有のガードレールや意思決定分類の導入へと向かっています。また、機密データの主権を確実に維持するために、ローカルファーストのソリューションを構築し、セキュリティフレームワークを再構成しています。さらに、企業での導入において監査可能性と説明可能性を必須条件とし、さらには、従来のプライバシーが次世代にとって重要であるかどうかという、より根本的な問いについても検討が進められています。

結局のところ、2026年の基盤プロジェクトであるということは、より大きな責任を伴います。コミュニティを重視し、差し迫った重要な課題に焦点を当て、相互運用性と標準化を確保することで、業界はAIが人間の主体性を奪うものではなく、人間の能力を高めるためのツールであり続けることを保証できるのです.

以下の提言は、4回のパネルディスカッションからまとめられたものです：

1. 説明責任と法的枠組みの確立

- エージェントによる行動に対する人間の責任を定義する、明確な法的基準と枠組みを確立します。特に、エージェントが自律的に行動し、違法となる可能性のある悪意のある、あるいは意図せず有害な行為を行った場合に備えます。
- AIの理解を深め、リスクの責任の所在を明確にするため、企業ごとに特化した経営幹部向け教育を確実に実施します。これにより、経営陣や内部監査チームが、AIリスクがどのように軽減されているかを理解できるようにします。
- 規制産業におけるコンプライアンスを確保するため、人間の役割を、判断を下し、AI主導の意思決定に対して最終的な品質保証を行う方向へと転換する。

2. アイデンティティと意思決定の根拠の標準化

- エージェントのアイデンティティと信頼性に関する共通の用語集を構築し、業界を超えたコミュニケーションと標準化を促進します。
- AIの意思決定をその影響度に基づいて分類し、規制当局向けに監査可能な記録を提供する意思決定分類体系を開発します。

[6] Bonderud, Doug. AI decision-making: Where do businesses draw the line? IBM.
参照元: https://www.ibm.com/think/insights/ai-decision-making-where-do-businesses-draw-the-line

- エージェント向けのDNSに類似した、集中型または分散型のグローバルリポジトリを活用し、真の身元を証明するとともに、双方向のエージェント間取引を可能にします。

3. セキュリティとプライバシーの基盤を高度化

- 安全性確保のため、単なる最終出力だけでなく、内部の推論経路の監査を可能にするモデルやツールを導入します。
- エージェントと個人の間に信頼関係を構築するため、トランザクションごとにきめ細かなアクセス制御と権限のエスカレーションを行うためのガードレールを確立します。
- 主要なモデルが確立されたガードレールから逸脱しないよう監視する仕組みとして、「中間エージェント」を作成します。
- プライバシーやセキュリティを損なうことなく、エージェントが機密データを扱えるようにするオープンソースのレイヤーやサンドボックスについて連携します。

4. アイデンティティと意思決定の証拠を標準化

- メンテナーが、AIによって生成された大量のプルリクエストを自動的にスクリーニングおよびレビューするのに役立つツールやテンプレートを作成します。
- オープンソースプロジェクトが、プロプライエタリなモデルと競争力を維持するために必要な計算リソースと「トークン」を確保できるよう、資金調達メカニズムを整備します。
- 開発ライフサイクルの各段階で異なるエージェントを交換できるようにする、標準および機械可読な制御に重点を置きます。

注目のプロジェクトコミュニティ

Agentic AI Foundation

Agentic AI Foundation（AAIF）は、自律的で相互運用可能なAIシステムを実現する、急速に拡大しているエージェント型AI技術のエコシステムのためのオープンな財団です。MCP、goose、AGENTS.mdなどの創設プロジェクトを擁するAAIFは、エージェントがプラットフォームを超えて相互運用可能に動作するための中核となる標準とプロトコルを管理しています。透明性の高いガバナンスと幅広い業界の参加を通じて、AAIFはエージェント型AIの普及を推進し、そのインフラがオープンかつ予測可能な形で、実運用規模で進化することを保証しています。詳細については https://aaif.io/をご覧ください。

FINOS Fintech Open Source Foundation

The Fintech Open Source Foundation（FINOS）は、金融サービス分野におけるオープンソースソフトウェア、標準規格、および共同開発プラクティスの普及を促進することを使命とする非営利団体です。Linux Foundationの一員として、FINOSは、競合関係にある組織の開発者が、業務運営を変革する革新的なプロジェクトで協力できるよう、規制に準拠したプラットフォームを提供しています。大手金融機関、フィンテック企業、テクノロジーコンサルティング会社など100社以上の会員を擁するFINOSは、金融分野におけるオープンソースのイノベーションを牽引する最前線に立っています。

CLOUD NATIVE COMPUTING FOUNDATION

クラウドネイティブコンピューティングにより、組織はパブリッククラウド、プライベートクラウド、ハイブリッドクラウドにおいて、オープンソースのソフトウェアスタックを活用し、スケーラブルなアプリケーションを構築・運用できるようになります。Cloud Native Computing Foundation（CNCF）は、Kubernetes、Prometheus、Envoyなど、グローバルな技術インフラの重要なコンポーネントをホストしています。CNCFは、業界をリードする開発者、エンドユーザー、ベンダーを結集し、世界最大規模のオープンソース開発者カンファレンスを主催しています。世界最大のクラウドコンピューティング企業やソフトウェア企業、200社以上の革新的なスタートアップを含む約800のメンバーに支えられたCNCFは、非営利団体であるLinux Foundationの一員です。詳細については、 www.cncf.ioをご覧ください。

LF AI & DATA

Linux FoundationのプロジェクトであるLF AI & Data Foundationは、オープンソースのAI、データ、および分析プロジェクトの成長を促進し、持続的に支えています。世界をリードするテクノロジー企業の支援を受けているLF AI & Data は、AI開発におけるコラボレーションとイノベーションのための中立的な場を提供しています。詳細については、 https://lfaidata.foundationをご覧ください。

LF DECENTRALIZED TRUST

LF Decentralized Trustは、組織が安全で強靭なコードを用いてイノベーションを起こせるよう支援する技術のオープンな開発を行う、中立的な拠点です。これは、デジタルファースト経済に必要な透明性、信頼性、セキュリティ、効率性を提供する幅広い技術や標準規格を扱う、Linux Foundationの代表的な組織です。多様でグローバルなメンバーやコミュニティの支援を受け、LF Decentralized Trustは、ブロックチェーン、台帳、ID、暗号技術、および関連技術からなる拡大するエコシステム全体において、オープンソースのベストプラクティスを推進しています。詳細は、 www.lfdecentralizedtrust.orgをご覧ください。

PyTorch Foundation

PyTorch Foundationは、オープンソースのPyTorchフレームワークおよび幅広い革新的なオープンソースAIプロジェクトを支援する、コミュニティ主導のハブです。Linux Foundationが運営を主催するPyTorch Foundationは、モデルのトレーニングや推論から、特定分野向けのアプリケーションに至るまで、AIライフサイクル全体にわたるコラボレーションのための、ベンダー中立で信頼性の高い拠点を提供しています。オープンなガバナンス、戦略的な支援、そして世界中の貢献者コミュニティを通じて、PyTorch Foundationは、開発者、研究者、企業がAIを大規模に構築・展開できるよう支援しています。詳細は、 https://pytorch.org/foundationをご覧ください。

謝辞

著者一同は、フォーラムの開催ならびにイベント後の録画および速やかな文字起こしの提供を担当してくださったLF Eventsチームの皆様に感謝申し上げます。また、ファシリテーターのDaniela Barbosa、Hart Montgomery、Jono Bacon、Matt White、Gabriele Columbro、ならびに原稿のレビューとフィードバックをいただいたアドバイザーのMichael Dolan、Greg Kroah-Hartman、Alex Salkaverにも感謝いたします。最後に、PDFの作成を担当してくださったLinux Foundation Creative Supportチームの皆様にも感謝申し上げます。

著者について

HILARY CARTER は2021年にLinux Foundationに参画し、オープンソースの動向、機会、課題に関する実証的な知見を提供するために設立された「LF Research」の立ち上げと統括を担当しました。Linux Foundationに参画する前、Hilaryはトロントを拠点とするBlockchain Research Instituteのマネージング ディレクターを務めていました。同機関は、ブロックチェーン技術に特化したグローバルなシンジケート型シンクタンクです。彼女は、オープンソースのイノベーションおよび業界横断的な導入に焦点を当てた、200件近くの研究プロジェクトに貢献してきました。Hilaryはロンドン・スクール・オブ・エコノミクス（LSE）で経営学の修士号を取得しており、カナダとアイルランドの二重国籍を持っています。

ANNA HERMANSENは、Linux Foundationの上級研究員兼エコシステム・マネージャーを務めており、研究プロジェクトを主導するとともに、Linux Foundationの研究活動全般の管理を支援しています。彼女は、ヘルスデータインフラストラクチャおよびオープンソースAIに関する定性的研究や系統的レビュー研究を行っており、カンファレンスやワーキンググループでこれらの研究成果を発表してきました。彼女の関心は、AI、プレシジョン メディシン、および健康データの共有が交差する領域にあります。彼女はジェネラリストとして、クライアントサービス、プログラムの実施、プロジェクト管理、そして学術界、企業、ウェブユーザー向けの執筆経験を有しています。Linux Foundationに加入する前は、Blockchain Research InstituteとBC Cancer’s Research Instituteという2つの異なる研究プログラムで勤務していました。彼女はブリティッシュコロンビア大学にて、公衆衛生学の修士号および国際関係の学士号を取得しています。

THE LINUX FOUNDATION | Research

2021年に設立された Linux Foundation Research は、拡大を続けるオープンソースの協業の規模を調査し、新たな技術トレンド、ベストプラクティス、そしてオープンソースプロジェクトが世界にもたらす影響に関する洞察を提供しています。プロジェクトデータベースやネットワークを活用し、定量的・定性的手法におけるベストプラクティスを徹底することで、Linux Foundation Researchは、世界中の組織の利益のために、オープンソースに関する知見の頼れる情報源を構築しています。

本論文を引用される際は、以下の形式で記載してください：Hilary Carter および Anna Hermansen、「Open Source and the Future of AI: How Agents are Disrupting Our Systems, Our Precedent, and the Human Role in Software」、The Linux Foundation、2026年4月。

この日本語文書は、英語版を機械翻訳し、「Open Source and the Future of AI: How Agents are Disrupting Our Systems, Our Precedent, and the Human Role in Software」の参考訳として、The Linux Foundation Japanが便宜上提供するものです。
翻訳協力：吉田行男

facebook.com/TheLinuxFoundation
x.com/linuxfoundation
linkedin.com/company/TheLinuxFoundation